Disentangling the mechanisms underlying infant fixation durations in scene perception

Running head: Modeling infant fixation durations in complex scenes

Disentangling the mechanisms underlying infant fixation durations in scene perception: A computational account

Irati R. Saez de Urabain[1], Antje Nuthmann[2], Mark H. Johnson[1], Tim J. Smith[1,*]

[1]Centre for Brain and Cognitive Development, Birkbeck College, University of London, UK

[2] School of Philosophy, Psychology and Language Sciences, Psychology Department, University of Edinburgh, UK

*Corresponding author. *E-mail address:* tj.smith@bbk.ac.uk (T. J. Smith)

Abstract

The goal of this article is to investigate the unexplored mechanisms underlying the development of saccadic control in infancy by determining the generalizability and potential limitations of extending the CRISP theoretical framework and computational model of fixation durations (FDs) in adult scene-viewing to infants. The CRISP model was used to investigate the underlying mechanisms modulating FDs in 6-month-olds by applying the model to empirical eye-movement data gathered from groups of infants and adults during free-viewing of naturalistic and semi-naturalistic videos. Participants also performed a gap-overlap task to measure their disengagement abilities. Results confirmed the CRISP model's applicability to infant data. Specifically, model simulations support the view that infant saccade programming is completed in two stages: an initial labile stage, followed by a non-labile stage. Moreover, results from the empirical data and simulation studies highlighted the influence of the material viewed on the FD distributions in infants and adults, as well as the impact that the developmental state of the oculomotor system can have on saccade programming and execution at 6 months. The present work suggests that infant FDs reflect on-line perceptual and cognitive activity in a similar way to adults, but that the individual developmental state of the oculomotor system affects this relationship at 6 months. Furthermore, computational modeling filled the gaps of psychophysical studies and allowed the effects of these two factors on FDs to be simulated in infant data providing greater insights into the development of oculomotor and attentional control than can be gained from behavioral results alone.

Keywords. Infancy, fixation durations, scene perception, saccadic control, computational modeling.

1. Introduction

From the moment we are born most daily activities involve constant decisions about where and when to move our eyes next. During active visual sampling our eyes may remain stable at a point (fixations, during which visual encoding occurs) or perform fast ballistic movements (saccadic eye movements; Matin, 1974; Ross, Morrone, Goldberg, & Burr, 2001).

In infants, FDs have been associated with the developmental state of the oculomotor system (Bronson, 1994; Johnson, 1990) and with visual and cognitive processes such as attention, information processing, memory or anticipation (e.g., Harris, Hainline, Abramov, Lemerise, & Camenzuli, 1988; Papageorgiou et al., 2014). Nevertheless, due to practical and technical limitations in testing young infants not much is known about the development of the mechanisms underlying FDs and saccade programming.

Computational modeling allows us to describe, predict and explain data that is itself unobservable (Lewandowsky & Farrell, 2011). This method is particularly useful in developmental science, where it permits the investigation of aspects of eye-movement control that could not be analyzed otherwise (Reichle et al., 2013). The present article aims to utilize a theoretical framework and computational model of FDs in scenes and determine its generalizability to infants (CRISP, a timer (C)ontrolled (R)andom-walk with (I)nhibition for (S)accade (P)lanning model; Nuthmann & Henderson, 2012; Nuthmann, Smith, Engbert, & Henderson, 2010), in order to investigate the mechanisms underlying FDs in 6-month-olds. Notably, the model assumes that saccades are programmed in two stages, an initial labile phase during which saccade programs can be altered or cancelled and a subsequent non-labile phase in which programs cannot be cancelled. Whether infant saccadic programming operates via these two phases is not known. In addition, the CRISP model will be used to examine whether FDs at

this age are affected by developmental aspects of the oculomotor system and/or by visual and cognitive processing. For this purpose, we report fixation-duration data from 6-month-old infants and adults who each viewed dynamic scenes, and two simulation studies that will test whether the data from both infants and adults can be explained by a single model architecture, with age-specific and task-specific influences realized by differences in parameter settings. In the following sections, we review the past research on FDs in adults and infants, and introduce the background literature on modeling FDs with CRISP.

1.1 FDs and saccade latency in adults

The relationship between FDs and visual and cognitive processing has been extensively investigated in skilled adult reading (Rayner, 1998, for review) and, more recently, also in the context of scene viewing (Nuthmann, 2016, for review). For instance, factors such as the viewing task (search vs. memorization; Castelhano, Mack, & Henderson, 2009; Nuthmann et al., 2010), the visual characteristics of the stimulus (e.g., luminance, image degradation; Loftus, 1985; Walshe & Nuthmann, 2014), the semantics of the scene (e.g., Henderson, Weeks, & Hollingworth, 1999; Loftus & Mackworth, 1978; Wu, Wick, & Pomplun, 2014), or familiarity (e.g., Althoff & Cohen, 1999) can affect gaze control and FDs. Collectively, these findings demonstrate that visual and cognitive processing demands are associated with differences in FDs.

The mechanisms underlying saccadic control in adults have been greatly investigated by studying saccadic responses in simple saccade-targeting tasks. In the double-step paradigm, for instance, participants are instructed to follow a target while it makes two successive movements or steps that are separated by a varying temporal gap (Becker & Jürgens, 1979; Camalier et al., 2007; Findlay & Harris, 1984; Westheimer, 1954). Findings from double-step studies have

provided evidence for parallel programming of saccades in which saccade programming occurs in two stages: an initial labile stage which is subject to cancellation, followed by a non-labile stage that cannot be cancelled. Becker and Jürgens (1979) showed that participants' performance is best predicted by the time *D* elapsing between the onset of the second target step and the onset of the first saccade. If *D* is short (< 70 ms), the response saccade is directed to the first target location. The saccade program to this location was already fully specified; in other words, it was in its non-labile stage of development when it could no longer be altered. As *D* increases, an amplitude transition function emerges, with the first saccade landing progressively closer to the second, final location of the target. When the first saccade lands on the second target, the oculomotor system began programming a saccade to the second target location while the saccade program for the first target location was still in its labile stage of development. In this situation, the first program is cancelled and only the second program is executed, prolonging the saccade latency and hence the duration of the fixation. Recently, Walshe and Nuthmann (2015) adopted the double-step paradigm to a scene-viewing context and showed that saccade cancellation processes generalize to scene viewing, and that cancelling a saccade prolongs FDs. The general finding that saccades are programmed in two stages has been adopted in computational models of fixation behavior in reading (Engbert, Longtin, & Kliegl, 2002; Reichle, Pollatsek, Fisher, & Rayner, 1998) and scene viewing (Nuthmann et al., 2010).

In infant research, video-based double-step paradigms have been used to investigate developmental changes in spatial remapping of saccade trajectories across the two saccades, but the paradigm has yet to be used to isolate the timing of the transition from labile to non-labile stage of saccade programming in infants (Brown et al., 2003; Gilmore & Johnson, 1997a, 1997b).

1.2 FDs and visual and cognitive processing in infancy

The mechanisms that control FDs in infants remain poorly understood. Evidence suggests that as early as 3- to 4-month-old infants' looking behavior (e.g., looking times to a particular stimulus) can be influenced by cognitive factors related to the visual input such as expectations of spatiotemporal object continuity and causality (e.g., Leslie & Keeble, 1987; Spelke, 1990). However, such studies examine rather coarse shifts in attention (look at the scene vs. don't look) with relatively few studies paying attention to the micro-dynamics of visual and cognitive processing (such as FDs).

For instance, some studies have investigated infants' scanning abilities and FDs when presented with familiar and non-familiar complex dynamic stimuli. Hunnius & Geuze (2004) followed infants between the ages of 6 and 26 weeks and presented them with a video of their mother's face, and an abstract video. They found that infants only adapted their eye-movements according to the type of stimulus from 14 weeks on, showing longer mean FDs for the abstract unfamiliar condition. Additionally, the median fixation duration did not stabilize before 18 weeks, which is slightly later than what has been reported for static stimuli (Bronson, 1990). These findings suggest that FDs in infancy can also reflect the visual and cognitive processing of the visual input, even though it is still unclear whether these factors have the same influence in infants and adults.

1.3 Neural mechanisms underlying eye-movement control in infancy

Whilst the subcortical structures involved in saccadic generation are relatively developed at birth (e.g., superior colliculus), cortical pathways associated with the generation of more complex eye-movements (e.g., the frontal eye fields, FEF) remain underdeveloped until 3 to 4 months of age (e.g., Atkinson, 2000; Bronson, 1974; Johnson, 1990, 2011). At around 1 month

postnatal age, unregulated tonic inhibition of the superior colliculus prevents infants from consistently moving their eyes from a point of foveation. This phenomenon is commonly known as 'sticky fixation' or 'obligatory attention' (Atkinson, 2000; Braddick et al., 1992; Farroni, Simion, Umiltà, & Barba, 1999; Frick, Colombo, & Saxon, 1999; Johnson, 2011) and is thought to diminish from 3 to 4 months with the increasing cortical control over saccades. Sticky fixation is thought to occur due to problems with "disengagement", defined as the difficulty in generating an eye-movement after a fixation (Johnson, 1990).

The ability to disengage from a central target to shift the gaze to a peripheral target has traditionally been evaluated using the gap-overlap paradigm (Atkinson, Hood, Wattam-Bell, & Braddick, 1992; Farroni et al., 1999; Hood & Atkinson, 1990, 1993; Johnson, Posner, & Rothbart, 1991). Reaction times are usually faster on gap trials, where the central target disappears and after a temporal gap (e.g., 200 ms) the peripheral target appears. During overlap trials, the central target stays on after the peripheral target appears. Sometimes, baseline trials are additionally included, in which the peripheral target appears without a temporal gap after the central target disappears (e.g., Elsabbagh et al., 2009; Wass, Porayska-Pomsta, & Johnson, 2011). In this case, disengagement latencies can be calculated by subtracting the baseline latencies from the overlap latencies.

In infancy research, longer disengagement latencies have been associated with greater immaturity of the visual system (e.g., Butcher et al., 2000; Matsuzawa & Shimojo, 1997), particularly during the first 6 months of life, when the neurological structures involved are thought to develop rapidly and approach their adult form (Rothbart, Posner, & Rosicky, 1994). Nevertheless, evidence from various neurophysiological and behavioral studies suggests that the neural mechanisms involved in the execution of saccades are still undergoing development, or

need to increase in efficiency, by 6 months postnatal age, despite hundreds of thousands of practice saccades (e.g., Butcher et al., 2000; Hood & Atkinson, 1993; Matsuzawa & Shimojo, 1997).

1.4 Modeling FDs

Empirical studies alone may be insufficient to answer questions related to the extent to which longer FDs reflect a less developed oculomotor system, or differences in the speed of visual and cognitive processing. Most current models of visual attention have focused on explaining where viewers fixate (Borji & Itti, 2013, for review). In comparison, only a few models provide control mechanisms for the duration of fixations to shed light on the mechanisms underlying saccade generation and the encoding of visual information. Most of these models were developed for the task of reading, such as the E-Z Reader model (Reichle et al., 1998; Reichle, Rayner, & Pollatsek, 2003) and the SWIFT model of saccade generation during reading (Engbert et al., 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005). Both models implement the notion of saccade programming being completed in two stages, as suggested by results from double-step experiments (Becker & Jürgens, 1979).

The CRISP model is a computational model and a theoretical framework that accounts for FDs in adult scene viewing (Nuthmann et al., 2010; see Figure 1 for a model overview). The model architecture can be summarized with three main assumptions on saccade timing and programming: (1) the accumulation of activity to a saccade threshold is implemented via a random-walk process and is responsible for generating inter-saccadic intervals and hence variations in FDs; (2) saccade programming occurs in two stages: an initial, labile stage that is subject to cancellation, and a later, non-labile stage; and (3) processing difficulty can inhibit saccade timing and programming in a moment-to-moment fashion. For the latter, the default

mechanism is that increased processing demands slow down the random-walk saccade timer, which delays the initiation of a new saccade program and eventually leads to longer FDs. Moreover, processing difficulties can even cancel an ongoing labile saccade program (red cross in Figure 1), which extends the duration of the current fixation; this specific mechanism was implemented in model simulations that investigated to what extent fixation durations were controlled directly and in real time by the current scene image (Nuthmann et al., 2010).

The CRISP model combines autonomous timing and temporally overlapping saccade programming, a feature it shares with the SWIFT model (Engbert et al., 2005) and the ICAT model (Trukenbrod & Engbert, 2014). To reiterate, results from the double-step paradigm suggest that saccade programming is completed in different stages and that saccade programs can partially overlap in time (Becker & Jürgens, 1979; Walshe & Nuthmann, 2015). In the CRISP model, there are two different cases of temporally overlapping saccade programs. It can happen that the saccade timer initiates a new saccade program while another one is still active but has already progressed to the non-labile stage of development. In this case, both saccade programs will be executed; typically, this will result into a rather short fixation between the two saccades. More importantly, if a new saccade program is initiated while another program is still in its labile stage of development, the first program will be cancelled and only the second one will be executed. When a saccade cancellation occurs within a fixation, the fixation duration tends to be longer than without it (Nuthmann et al., 2010).

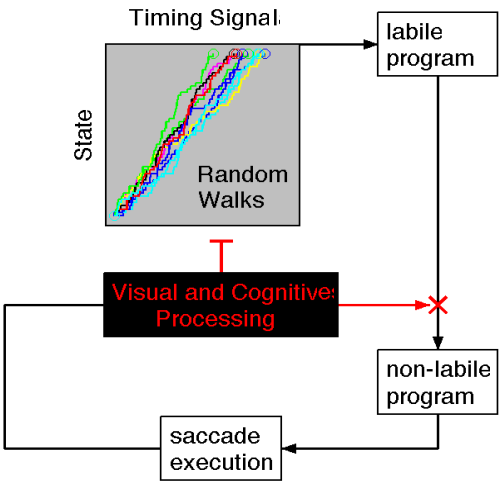

Figure 1. Overview of the CRISP model. A random-walk timing signal accumulates toward a threshold. Accumulation rate is affected by ongoing visual and cognitive processing in a moment-to-moment fashion. Once the threshold is reached, a new saccade program is initiated. The saccade program first enters a labile stage, which is subject to cancellation if another saccadic program is subsequently initiated. At the end of the labile stage, a point of no return is reached. During the following non-labile stage, the saccade program can no longer be cancelled. Finally, the saccade is executed. Fixation durations are the time intervals between successive saccades. Figure from Nuthmann et al. (2010).

Mathematically, in its basic form the CRISP model comprises five parameters related to saccade timing and programming (t_{sac}, N, τ_{lab}, τ_{nlab}, $\tau_{ex;}$ see below for definitions). The random-walk timing signal accumulates toward a threshold, and once the threshold is reached, a new saccade program is initiated. The transition rate for the random-walk's increments determines how fast the process of saccade timing operates. The transition rate r_1 is defined as

Equation 1

$$r_1 = \frac{N}{t_{sac}}$$

where t_{sac} is the mean duration of the timing signal (i.e., the mean time interval between two commands to initiate a saccade program). N denotes the number of states the random-walk process can adopt and determines the variance of the timing signal; a low N leads to high

variance and thus broad fixation duration distributions. The random-walk is implemented as a discrete-state continuous-time Markov process with exponentially distributed waiting times between elementary transitions (for further details, see Nuthmann et al., 2010).

Furthermore, saccade latency l_{sac} is derived as the sum of the implemented saccade programming stages, i.e.,

Equation 2

$$l_{sac} = \tau_{lab} + \tau_{nlab}$$

where τ_{lab} and τ_{nlab} denote the average duration of the labile and non-labile phases, respectively. At the end of the non-labile saccade programming stage, the saccade is executed (with average duration τ_{ex}).

For some of the considerations below, saccade programming parameters (τ_{lab}, τ_{nlab}) serve as an index for the development of the oculomotor system, and saccade timing parameters (t_{sac}, N) are interpreted as an indicator for the speed of visual-cognitive processing.

Saccade programming processes are assumed to be stochastic in nature. Therefore, for each realization of the model simulation, parameter values for the different saccade programming phases (τ_{lab}, τ_{nlab}, and τ_{ex}) are drawn from gamma distributions. This introduces unsystematic variability in the duration of saccade programming stages. In simulations with models of eye-movement control in visual-cognitive tasks like reading and scene viewing, the mean duration of saccade execution (τ_{ex}) is treated as a fixed parameter. Typically, modellers choose a value that is in general agreement with empirical data (e.g., Engbert et al., 2005; Reichle et al., 1998); sometimes, the specific value is directly estimated from the mean saccade duration in the empirical data that are to be modelled (Nuthmann et al., 2010). For the remaining

parameters, best-fitting values can be determined using an optimization technique; it is important to ensure that the range of possible values for each parameter is psychologically and/or neurophysiologically plausible. In its implementation, the model generates sequences of saccades and fixations, whereby FDs are the time intervals between successive saccades (excluding saccade execution).

1.5 The current study

The current investigation includes an empirical study and two simulation studies with the CRISP model (Nuthmann et al., 2010). Our main goal was to run simulations with the CRISP model to elucidate the unexplored mechanisms underlying oculomotor control in 6-month-olds. Moreover, we wanted to study differences between long and short lookers, based in the assumption that long lookers might present a less developed oculomotor system (see Papageorgiou et al., 2014; Papageorgiou, Farroni, Johnson, Smith, & Ronald, 2015).

For the empirical study, we recorded the eye movements of 6-month-old infants and adults. Participants were presented with two different complex dynamic viewing conditions (naturalistic and semi-naturalistic videos) and performed the gap-overlap paradigm. The free viewing tasks were used to analyze the micro-dynamics of visual and cognitive processing during spontaneous orienting by measuring FDs. The gap-overlap task was used to measure participants' disengagement abilities.

We predicted that the empirical study would show (1) differences in FDs across viewing conditions and age groups (infants vs. adults), and (2) a positive relationship between FDs and disengagement latencies in both age groups, particularly in infants due to the underdeveloped state of their oculomotor system. The goal of simulation study 1 was to (3) test the CRISP model's generalizability with data from 6-month-old infants by analyzing differences between

infants and adults through changes in the model's input parameters. If the model was unable to recover the specific features of fixation-duration distributions in infants (e.g. mode, skew, tail), this would suggest that some of the model assumptions do not readily apply to eye-movement control during infancy. Simulation study 2 compared simulations from long and short lookers to investigate whether (4) FDs in infancy are predominantly influenced by the developmental state of the oculomotor system, or by visual-cognitive processing speed, or by both to varying degrees. Given that difficulty in disengagement is commonly associated with poorer developmental states of the oculomotor system (Matsuzawa & Shimojo, 1997), we hypothesized that the absolute and relative values for the labile, and especially non-labile stages of saccade programming would be prolonged in infants with longer disengagement latencies.

2. Empirical study

2.1 Methods

2.1.1 Participants

In total, 18 typically developing 6-month-olds and 18 adults (10 female) with no known visual impairments participated in this experiment. Seven infants had to be excluded from the analysis due to data quality issues (n = 5) or fussiness (n = 2). Thus, the analysis included 11 infants (mean age = 171.45 days, range = 150 to 191 days, 6 girls). Most infants (n = 10) were of Caucasian middle socioeconomic status. The infants were recruited via magazine advertisements, social networking media and flyers. The study protocol was approved by the relevant Ethics Committee.

2.1.2 Apparatus

Participants' gaze was monitored using a Tobii TX300 eye-tracker running at a sampling rate of 120 Hz. This particular eye-tracker model tolerates large head movements, allowing the

infants to move naturally in front of the stimuli presentation screen. The stimuli were presented on a 23'' wide screen TFT monitor attached to the eye tracker unit by using a custom-made MATLAB program (MATLAB version R2010a 32-bit). This program made use of the T2T (Talk to Tobii) package and the Psychophysics Toolbox 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007). The sounds were played through stereo external speakers located at both sides of the screen. The participant was monitored and recorded through an external video camera located under the Tobii screen by using the ScreenFlow screen-casting software. The objects and figures for the gap-overlap paradigm were created using Adobe Photoshop CS6. The naturalistic videos were recorded with a full high-definition camera and assembled with Final Cut Pro X Version 10.0.5.

2.1.3 Dynamic stimuli

The spontaneous looking task includes two different viewing conditions: the naturalistic condition in which three people performed baby-friendly actions, either simultaneously or at different times; and the semi-naturalistic condition in which the people were substituted with simple geometrical shapes rotating either simultaneously or at different times. The two types of stimuli were presented in color and were approximately equal in size. The background music that accompanied the stimuli consisted of fragments of instrumental songs.

Each naturalistic video (see Figure 2) lasted between 20 and 25 seconds and comprised three regions of interest (right, left and center of the screen) in which three different volunteers performed a random baby-friendly action, e.g. waving a balloon or tossing a ball. The actors were filmed against a green screen, which allowed the background to be replaced with a baby-friendly static image using chromakeying. In the semi-naturalistic condition (see Figure 3) the characters were substituted with simple geometrical shapes (colorful triangles, rectangles,

circles, squares, pentagons, hexagons or ellipses) that were rotating clockwise simultaneously or at different times. The position of the right and the left elements corresponded to a visual angle of 18.26°. This design aims to uniformly distribute the semantic and motion information across the screen. The set included 10 different naturalistic videos and 12 semi-naturalistic videos. The two extra videos were added in order to compensate for a possible decrease in attention to this condition towards the end of the experiment, and to obtain the largest number of fixations possible.

Figure 2. Screenshot from one out of ten naturalistic videos used in the study.

Figure 3. Screenshot from one out of twelve semi-naturalistic videos used in the study.

2.1.4 Gap-overlap paradigm

In a gaze-contingent paradigm, experimental events are contingent on participants' online eye-movement behaviors, thereby permitting the study of eye movements in a truly interactive manner (e.g., McConkie & Rayner, 1975; Wass et al., 2011; Wilms et al., 2010). Our experimental design included two gaze-contingent tasks: a gap-overlap and a double-step task. However, there were not enough valid double-step trials per condition in order to identify the threshold between labile and non-labile stages; therefore, the results from the double step paradigm are not included in this article.

Each gap-overlap trial starts with the presentation of a central stimulus. In the moment the infant fixates this stimulus, the peripheral target will randomly appear either on the right or the left side of the screen (18°). Depending on the time at which the central stimulus disappears, a given trial will be flagged as overlap, baseline or gap (see Introduction). The gap trials evaluate

the facilitation effect emerged from a temporal gap of 200 ms preceding the peripheral presentation. When the participant looks at the peripheral stimulus, a short animation is presented. If the participant does not look at the peripheral stimulus after 4 seconds, the next trial starts. The disengagement latencies are calculated by subtracting the baseline condition from the overlap condition. The central stimulus was a colorful ball, while the peripheral one was a cloud.

A total number of 72 trials were presented. Forty percent of these trials were gap trials, 30% were overlap trials and the remaining 30% were baseline trials. They were presented in groups of 12 trials (iteration) and alternated with other tasks of the study.

2.1.5 Procedure

Participants were welcomed in a lab waiting room in which infants acclimated to the experimenter and the lab. Next, they were tested individually in a darkened room while sitting on a baby-chair located 60 cm away from the monitor. Prior to starting the experiment, the infants were calibrated with a five-point calibration. Infants were shown a moving puppet accompanied by a sound on each point of the screen and on the center until the target was fixated. This process was repeated until the infant looked at the five points accurately. Adults were seated on a chair located 60 cm away from the monitor. They also completed the infant five-point calibration procedure.

The different tasks of the study were alternated in the following order: (1) twelve gap-overlap trials, (2) one naturalistic video presentation (20-25 seconds), (3) twelve double-step trials, and (4) one semi-naturalistic video presentation (20-25 seconds). This process was repeated 6 times. Next, 6 additional iterations of naturalistic and semi-naturalistic videos were presented. In total, participants viewed 10 naturalistic videos, 12 semi-naturalistic videos, 72

gap-overlap trials, and 72 double-step trials. In case the infant became upset during the study the program was stopped.

2.2 Analysis

2.2.1 Cross-validation

2.2.1.1 Fixation Durations (FDs)

Eye-tracking data from infants may contain considerably higher levels of noise than data from more compliant participants such as adults due to various factors including their high degree of movement, lack of compliance to the task, poor calibration and corneal reflection disturbances due to the underdeveloped cornea and iris (Hessels, Andersson, Hooge, Nyström, & Kemner, 2015; Saez de Urabain, Johnson, & Smith, 2015; Wass, Smith, & Johnson, 2013). To account for this potential quality/age confound, dedicated in-house software for parsing and cleaning eye tracking data has been developed (GraFix, Saez de Urabain et al., 2015). This software allows valid fixations to be salvaged from low-quality datasets whilst also removing spurious invalid fixations. In the present study, both adult and infant datasets were parsed using GraFix's two-stage semi-automated process (see Appendix A for details). The second stage of GraFix involves manual checking of the fixations detected automatically during the first stage. This manual coding stage was validated by assessing the degree of agreement between two different raters. One rater was one of the authors (IRSdU). An external coder, naïve to the expected outcomes, was trained to code fixations from infants featuring low- and high-quality data. The coder had to (1) run GraFIX automatic detection algorithms using the parameters from Appendix A and to (2) manipulate the outcome in order to remove artifactual fixations or add those undetected by following the predefined guidelines. In total, the external coder coded 10 % of the data analyzed for this study. The inter-rater reliability between the mean FDs was

evaluated using the Intraclass Correlation Coefficient (ICC; Hallgren, 2012). A strong agreement was found (ICC = .898, p = .05).

2.2.1.2 Gap-overlap paradigm

All trials from the gap-overlap paradigm were inspected by one of the authors (IRSdU) to manually exclude those where (1) the accuracy was not good enough to trigger the gaze-contingencies, or (2) the infants looked away during the presentation of the peripheral stimulus. An external coder, who was naïve to the expected outcomes, evaluated the validity of the gap-overlap trials for 10% of the data. Interrater reliability was measured using the Kappa statistic to determine consistency among raters. This analysis showed a strong agreement between the two coders, Kappa = .878 (p < .001), 95% CI (.833, .923).

2.2.2 Data-quality analysis

Spatial precision is defined as the consistency in detecting and calculating gaze points accurately. Low spatial precision can seriously affect the detection of fixations and hence it is essential to measure and report it. There are a number of methods to calculate spatial precision, such as the RMS (commonly used by manufacturers) or the standard deviation, which measures the dispersion of each sample from a mean value (Holmqvist et al., 2011). The RMS was calculated for each participant.

In low-quality data it is less likely to find clean long fixations that can be reliably detected even if the data are hand-coded. For this reason, one could expect to find negative correlations between FDs and precision measures, where participants showing lower data quality would also feature proportionally shorter fixations. Correlational analysis between the RMS-s and FDs did not find any correlations for the group of 6-month-olds nor for the group of adults, confirming that data quality did not affect the experimental results in the present study.

2.2.3 Statistical analysis

The mean fixation duration for each participant and viewing condition was calculated after excluding all fixations with a duration greater than two standard deviations from the initial mean. Hence, fixations included in the analysis accounted for about 95% of the set for each participant. This procedure was necessary to exclude very long or very short fixations on a participant basis, rather than establishing a minimum and maximum fixation duration threshold for all participants together, regardless of their age. This criterion was established in order to exclude those rare very long fixations (e.g., 10 seconds) that were a consequence of the infant's sleepiness rather than a result of a cognitive or neural process. At the same time, this procedure ensured that long fixations, which are quite typical in infant populations, were not generally excluded.

For the group of infants, the gap-overlap trials with latencies shorter than 200 ms were excluded from the analysis, because such fast eye-movements are considered to be eye-tracking errors or anticipatory saccades (e.g., Canfield, Wilken, Schmerl, & Smith, 1995; Frick et al., 1999; Rose, Feldman, Jankowski, & Caro, 2002). In contrast, for the group of adults the minimum latency was lowered to 80 ms, since latencies below this value are considered to be anticipatory or express saccades (Fischer & Weber, 1993). Some studies also define a maximum latency limit (Elsabbagh et al., 2009; Johnson et al., 1991; Matsuzawa & Shimojo, 1997). However, given that our two participant groups are very different in age, one can expect to find group and inter-individual differences in their disengagement abilities and reaction times. For this reason, we decided to establish a conservative criterion by excluding all the trials that were one standard deviation above each participant's mean latency.

3. Simulation studies

3.1 Modeling infant FDs with the CRISP model

Only few computational models have attempted to explain cognitive development in infancy (for a review see Schlesinger & McMurray, 2012). Whilst some of these models have investigated different aspects of visual orienting in infancy (Carlson & Triesch, 2004; Schlesinger, Amso, & Johnson, 2007; Domsch et al., 2010; Sirois & Mareschal, 2002; for a review see Mareschal, 2010), to date no computational model has attempted to explain the specific mechanisms underlying saccadic control in infancy.

As outlined in the introduction, some evidence exists that infant FDs can be affected by developmental aspects of the oculomotor system and/or by visual-cognitive processing demands. As a result, one can expect to find more variability in FDs in infants compared to adults, or in other words, more positively skewed distributions of FDs (Harris et al., 1988). In order to model this variability in infant oculomotor control, the CRISP model (Nuthmann et al., 2010) was applied to both infant and adult data.

In the following, we revisit CRISP's theoretical assumptions and input parameters with the specific aim to explore whether they are, in principle, applicable to infant data and in what way age differences (infants vs. adults) could be represented by certain differences in parameter settings. From the perspective of age-specific influences on FDs, the following CRISP parameters contribute toward systematic differences in FDs: (1) the mean (t_{sac}) and variance (i.e., number of states N) of the random-walk timing signal, (2) the mean duration of the labile saccade program (τ_{lab}), and (3) the mean duration of the non-labile saccade program (τ_{nlab}). Moreover, cancelations of existing saccade programs contribute to the variability in fixation

durations, and they are particularly important for producing long tailed fixation duration distributions (see Nuthmann et al., 2010).

3.1.1 Random-walk timing signal

The CRISP model assumes that saccade programs are initiated according to some preferred idiosyncratic mean rate (Engbert et al., 2002) and based on this it implements a rhythmic saccade timer that is responsible for generating variations in FDs (Nuthmann et al., 2010). All living things are subject to spontaneous fluctuations of physiological and behavioral processes that maintain complex organisms in a fairly stable state (Wolff, 1991). Further, there is some evidence that the central nervous system as well as the oculomotor system are rhythmic in nature (McAuley, Rothwell, & Marsden, 1999). The temporal organization of spontaneous movements has also been studied in newborn infants, where it was found that the fluctuations in movement over time were not random but rather rhythmic (Robertson, 1982). The same can be interpreted when investigating saccadic eye-movements during the first month of life, which are believed to be mainly under subcortical control and hence are rapid and input-driven (Atkinson, 2000; Johnson, 1995, 2011).

3.1.2 Random-walk timing and speed of processing

In CRISP, decreases in processing speed and increases in processing difficulty will slow down the random-walk saccade timer. This delays the initiation of a new saccade program, which in turn leads to longer FDs. For our simulations, it is reasonable to assume that the mean and variance of the random-walk process underlying the initiation of a new saccade program are different for infants compared to adults. Specifically, CRISP would capture a slower speed of visual-cognitive processing in infants by a higher mean value t_{sac} for the random-walk timing signal. In addition, following previous viewing-task simulations with the CRISP model (scene

memorization vs. search: Nuthmann et al., 2010; scene viewing vs. reading: Nuthmann & Henderson, 2012) we allow both parameters of saccade timing (t_{sac} and N) to vary across viewing conditions. In sum, we predict that these values will vary across viewing conditions and age groups.

3.1.3 Two-stage saccade programming

In CRISP, once the random-walk process reaches threshold, a new saccade program is initiated. Saccade programming is completed in two stages: an initial, labile stage that is subject to cancellation and an ensuing, non-labile stage in which the program can no longer be cancelled. As outlined above, CRISP introduces stochastic (i.e., unsystematic) variability in the durations of the labile and non-labile stages in that, for each simulated saccade program, individual durations are drawn from gamma distributions with means τ_{lab} and τ_{nlab}, respectively. The question arises whether there are additional systematic age-dependent and/or viewing condition-dependent differences in the duration of saccade programming stages.

Investigating the two-step notion of saccade programming in infants using the double-step paradigm is challenging due to the need for a large number of trials in order to identify the transition from labile to non-labile stages of saccade programming. Prior infant studies using the double-step paradigm have instead focused on investigating developmental changes in spatial reference frames by observing where the saccades land rather than when they are launched (Brown et al., 2003; Gilmore & Johnson, 1997a, 1997b). However, infants' ability to inhibit a saccade has been investigated by using the anti-saccade paradigm (e.g., Johnson, 1995). In this task, after the appearance of a central stimulus a cue appears either on the left or right side. The infant is encouraged to look in the opposite direction by subsequently presenting a more attractive object at this location. Johnson (1995) found that by 4 months, infants were able to,

after a training period, reliably inhibit the saccade to the cue and move their eyes to the second more attractive target instead. Interestingly, the ability to inhibit a saccade programming concurs with the major development of the premotor areas of the frontal lobes – which contain the frontal eye fields – that occurs from 3 to 4 months. These findings suggest that saccade cancellation and hence the two stages of saccade programming may be present already by 4 months. Thus, assuming that infants may not be able to systematically cancel a saccade before 4 months, one could expect the non-labile program to be longer than the labile program before this age. From 4 months onwards, we predict (1) a relative increase of the labile stage with respect to the non-labile stage program, and (2) a gradual decrease in the absolute times for both labile and non-labile programs as the infant's saccadic control increases in efficiency. Furthermore, research has shown that even by 6 months, infant saccadic control is not as efficient as in adult participants (Butcher et al., 2000; Csibra, Tucker, & Johnson, 2001), thus both labile and non-labile stages are likely to be longer at this age compared to adults.

Under the assumption that both the existence and duration of the labile (τ_{lab}) and non-labile (τ_{nlab}) stages of saccade programing are subject to developmental changes in the oculomotor system, we treat them as an indicator of the system's developmental state. Longer labile and non-labile stages will, on average, generate longer FDs.

There is evidence suggesting that saccade programming and hence saccade latencies could be affected by various exogenous factors such as those observed in the gap-overlap paradigm, or endogenous factors such as those emerging from certain task instructions (e.g., Castelhano, Mack, & Henderson, 2009; Nuthmann et al., 2010). However, for reasons of model parsimony we assumed that saccade programming parameters do not vary across viewing conditions within a given age group.

3.2 Simulation study 1: Baseline

The goal of the first simulation was to test the CRISP model's generalizability to fixation-duration data from 6-month-old infants. Furthermore, we analyzed the age differences and the influence of the material viewed on FDs during scene viewing in both infants and adults.

3.2.1 Behavioral data

The empirical data set comprised 12315 adult fixations ($N_{naturalistic}$ = 5949; $N_{semi\text{-}naturalistic}$ = 6366) and 2675 infant fixations ($N_{naturalistic}$ = 1447; $N_{semi\text{-}naturalistic}$ = 1228). A mixed ANOVA exploring the differences in mean FDs with factors Viewing Condition and Age revealed a significant main effect of Age ($F(1,28) = 91.56$, $p < .001$) and Viewing Condition ($F(1,28) = 22.63$, $p < .001$), but no significant interaction ($F < 1$). The Age and Condition effects can clearly be seen in the group means (Figure 4). The adults presented significantly shorter mean FDs (naturalistic: M = 327 ms, SD = .045; semi-naturalistic: M = 377 ms, SD = 69) than the group of 6-month-olds (naturalistic: M = 644 ms, SD = .142; semi-naturalistic: M = 724 ms, SD = .120) and FDs were longer in both groups for the semi-naturalistic videos compared to the naturalistic videos (mean difference in adults = 50 ms, and in infants = 80 ms). Participants showing short or long mean FDs in one viewing condition also showed short or long mean FDs in the other condition, indicating stable individual differences in both 6-month-olds ($r(11) = .882$, $p < .001$) and adults ($r(18) = .615$, $p = .007$). In sum, while oculomotor control in 6-month old infants may not have reached adult levels, their FDs are already influenced by the viewing condition.

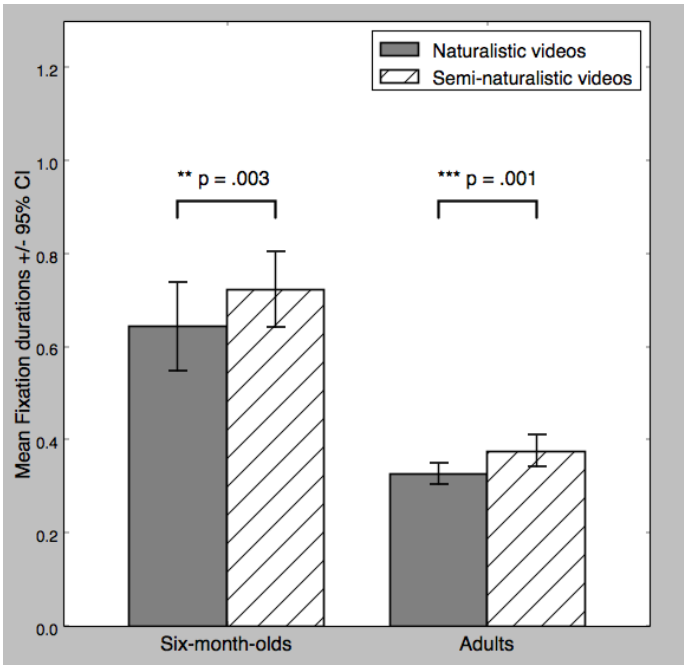

Figure 4. Mean fixation durations across viewing conditions for infants and adults.

3.2.2 Model adjustments

The goal of the simulations was to test the CRISP model's generalizability to infant data. To this end, the same model architecture was applied to both the infant and adult data. Thus, model generalizability was analyzed in the restricted sense of parameter changes, which is a more stringent test than adding new parameters to the model (Nuthmann & Engbert, 2009). Based on the theoretical and empirical considerations outlined above, it was decided which model parameters were allowed to vary across age groups and/or viewing conditions.

As explained earlier, the mean (t_{sac}) and variance (N, number of states) of the random-walk timing signal define the transition rate (r_1, see Equation 1) for the random-walk process, which determines how fast the process of saccade timing operates. For the present simulations we allowed the parameters t_{sac} and N to vary across different viewing conditions and age groups.

Parameter estimates for the different conditions serve as an index for the speed of visual and cognitive processing.

When the model simulates a given saccade program, the durations of the labile, non-labile and execution stages are drawn from gamma distributions (see also Engbert et al., 2005; Reichle et al., 1998). For example, the actual time required to complete the labile saccadic programming stage is sampled from a gamma distribution with $\mu = \tau_{lab}$ and $\sigma = \sigma_\gamma \times \mu$. In previous simulations of adult fixation durations in different scene-viewing tasks (Nuthmann et al., 2010), the means of the gamma distributions (μ) were either free or fixed parameters, whereas the relation between the standard deviation and mean (σ_γ) was fixed (e.g., at 0.33 or 0.25, Nuthmann et al., 2010). In the present simulations, the means for the labile and non-labile stages are free parameters, which we allow to vary across age groups while they are constant for viewing conditions within a given age group. The parameter for the standard deviation of the gamma distributions (σ_γ) is a fixed parameter. To accommodate the higher variability generally observed in infant data compared to adult data, it was set to 0.33 for the infant data and 0.25 for the adult data. These values were adopted from previous model simulations (Engbert et al., 2005; Nuthmann et al., 2010; Reichle et al., 1998, 2003). To reiterate, the variability σ of the gamma distributions is driven by both μ and σ_γ. It is important to note that, as the mean durations of stochastic processes increase, so too does their overall variability. For example, a higher mean duration for the labile stage of saccade programming is associated with a higher overall variability.

The mean duration of saccade execution (τ_{ex}) is a fixed parameter. For adults, we adopted a parameter value used in previous CRISP simulations (τ_{ex} = 37 ms, simulation study 2 in Nuthmann et al., 2010). For the infant simulations, we slightly increased this parameter value (τ_{ex}

= 50 ms), thereby acknowledging that saccade execution may be slower in infants than in adults (cf. Hainline, Turkel, Abramov, Lemerise, & Harris, 1984).

3.2.3 Modeling results

Free parameters were estimated using a genetic algorithm optimization technique that was used in previous simulation studies with the CRISP model (Nuthmann & Henderson, 2012; simulation study 2 in Nuthmann et al., 2010). For each age group, a total of six free parameters were fit: the mean durations for the labile and non-labile stages of programming; the mean and variance of the random-walk timing signal for the naturalistic condition; and the mean and variance of the random-walk timing signal for the semi-naturalistic condition. The ranges for estimating these values were based on previous oculomotor research from infants and adults; note that we used relatively wide ranges (especially for the group of adults). The genetic algorithm estimated the input parameters by minimizing a goodness-of-fit measure and by evaluating how much the simulated data differed from the empirical data. Note, the empirical data consisted of all fixations within an age group and viewing condition, collapsed together into a single fixation duration distribution. Individuals were not preserved due to the low number of fixations per participant and instead four distributions were modeled: 6 months naturalistic, 6 months semi-naturalistic, Adult naturalistic, and Adult semi-naturalistic. The parameter values generating the simulations that best resembled the empirical distributions are described in Table 1.

The goodness-of-fit measure (E) was calculated by assessing how much the simulated data diverged from the empirical data. For this purpose FD distributions and mean FDs were compared and the errors were added together (see Nuthmann et al., 2010, Appendix B). As a result of the long tail and the high skewedness of the fixation duration distributions presented in

this article, the simulation with the lower E did not necessarily indicate the best fit. Therefore, the parameters displaying the best fit were selected following the procedure described in Appendix B. Additionally, we report the mean-root-square error of the cumulative distributions (E_c). This function, which gives more weight to the peak of the distribution when calculating the error, constitutes an additional source for evaluating the disparity between distributions.

The results of the simulations are summarized in Figure 5. The two right panels show the data from adult observers, which are captured very well by the model simulations. Specifically, the simulated data reproduced the characteristic positive skew in fixation duration distributions. Compared to the adult data, the fixation duration distributions for the 6 month-olds are much noisier, partly due to the fewer number of fixations. For the infant data, the modal portion of the distribution is clearly shifted towards longer FDs, and the tail of the distribution is increased. These overall trends are captured well by the simulated data. Table 1 displays the best-fitting parameters for adults and infants.

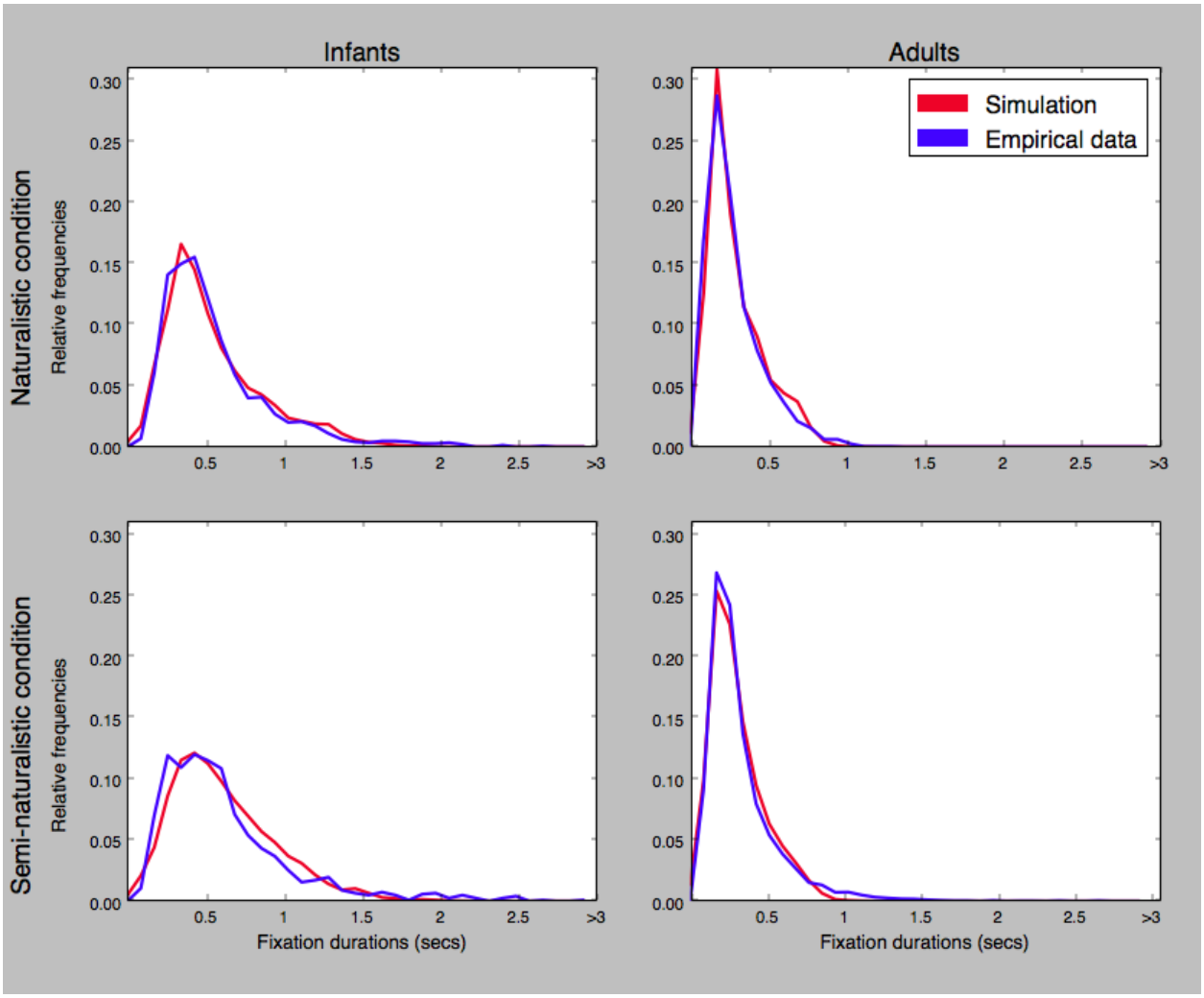

Figure 5. Simulation study 1: Simulated versus empirical fixation duration distributions. Data from 6-month old infants (left) are compared with data from adults (right) during viewing of naturalistic (top) and semi-naturalistic (bottom) videos.

[INSERT TABLE 1]

The simulation results suggest differences between age groups in both saccade timing and saccade programming parameters, with the infant group displaying considerably higher values (see Table 1). With respect to saccade timing parameters, the mean values for the random-walk timer from the infants group (naturalistic = 458 ms; semi-naturalistic = 496 ms) almost doubles the values for adults (naturalistic = 253 ms; semi-naturalistic = 271 ms). The saccade timing

variance (number of steps for the random-walk, N), on the other hand, does not vary much across age groups.[1] For both viewing conditions, the random-walk transition rate (N/ t_{sac}) is lower for infants than for adults. In CRISP, this transition rate can be modulated by visual-cognitive processing (Nuthmann et al., 2010). Therefore, the striking differences in saccade timing parameters between infants and adults may be interpreted such that both saccade timing and visual-cognitive processing speed are considerably slower for 6-month-olds than for adults, which manifests in longer fixation durations. Additionally, the mean duration of the random-walk timing signal and its transition rate were lower for semi-naturalistic stimuli than for naturalistic stimuli, which contributes to the observed differences in fixation durations with shorter durations for semi-naturalistic movies.

Pertaining to saccade programming parameters, the simulation results can be summarized as follows. First, there was no relative increase in the duration of the labile stage with respect to the non-labile stage when comparing infants and adults (the labile stage takes up 86.13% of saccade programming time in infants and 85.82% in adults). Second, there was a gradual decrease with age in the absolute durations of both labile and non-labile saccade programming stages (see Table 1). A possible interpretation of these results is that two-stage saccade programming may be fully functioning at 6 months, but it must develop further in order to reach adult-like levels.

[1] Three out of four estimates of the saccade timing variance (N) were simulated at the extremes of the possible value range (i.e. 5-20). This should not be seen as a limitation of the simulations as confirmatory simulations using broader ranges for N revealed a very similar pattern of estimates and fits for the empirical data. Moreover, N is only one of several parameters that determine the overall variability in the simulated fixation durations.

The main conclusion of this simulation study is that the CRISP model's architecture can be applied to infant data. However, it became apparent that the model (in its original form) has difficulty in dealing with the very long fixations common in infants (> 2 s). The simulation results suggest that the fixation duration data from 6-month-old infants were well fit by a model in which saccade programming progresses from a labile to a non-labile stage, suggesting that two-saccade programming may be active at this age. Furthermore, the CRISP model was able to capture the differences in FDs that resulted from the presentation of different viewing conditions in both age groups, suggesting that the influence of visual-cognitive processing in FDs can be reproduced by adapting the random-walk timer of the saccade generator.

Regarding age differences, the discrepancies in both saccade programming values (labile and non-labile stages) and saccade timing values (mean of the random-walk timing signal) between infants and adults are compelling, suggesting that at 6 months, further development of the oculomotor system is required to reach adult levels.

3.3 Simulation study 2: Modeling participants showing long and short disengagement latencies

This study aims to compare the differences in FDs between infants showing increased disengagement difficulties and those whose latencies were closer to adults.

3.3.1 Behavioral data

The gap-overlap paradigm was used to measure participants' disengagement abilities. Disengagement was calculated for each participant by subtracting the baseline mean latency from the overlap mean latency. Both infants and adults showed a clear gap effect such that overlap latencies were significantly longer than baseline latencies (6-month-olds: mean difference = 0.249 s, $t(10) = -4.795$, $p < 001$; adults: mean difference = 0.105 s, $t(17) = -6.004$, $p < .001$). For 6-month-olds, high correlations were found between disengagement latencies and

both naturalistic ($r(11) = .836$, $p = .001$) and semi-naturalistic FDs ($r(11) = .851$, $p = .001$). For adults, there was a systematic relationship between disengagement latencies and FDs in the naturalistic viewing condition ($r(18) = .471$, $p = .049$), but not for the semi-naturalistic condition ($r(18) = .251$, $p = .315$; see Figure 6). The same pattern of results was found when mean FDs for naturalistic and semi-naturalistic stimuli were log-transformed to achieve a near-normal distribution of the dependent variable. These results indicate that individual differences in the ability to disengage from a focal point have an influence on the timing of fixations. To some degree, this influence was also present in adults.

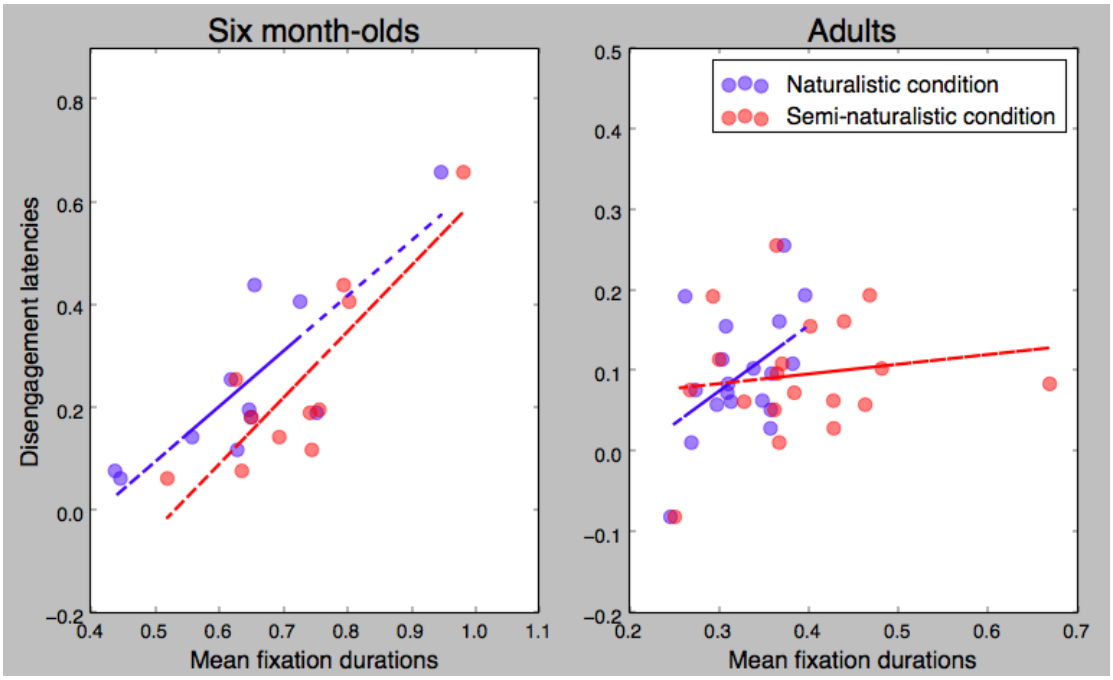

Figure 6. Correlations between disengagement latencies and fixation durations (in seconds) in infants and adults.

The model simulations were restricted to the infant data, with the aim to compare data from infants with particularly large or small disengagement difficulties, as a proxy for

oculomotor programming maturity. Due to the correlations between disengagement latencies and mean fixation durations (especially in the infants) participant mean fixation durations could have also been used to perform the same stratification but the decision was taken to use the disengagement latency as this is a more established developmental marker (Matsuzawa & Shimojo, 1997). Gathering a large amount of data from the same infant entails some technical and practical difficulties (Saez de Urabain et al., 2015). Due to the low number of fixations recorded from each infant, it was not possible to directly compare the two infants with the most extreme disengagement latencies (i.e. long and short). Data from at least three participants was needed to obtain an acceptable number of fixations for the modelling (> 200 fixations per condition). Therefore, the long disengagement (LongD) group included the three infants with the longest disengagement latencies (all of them above the mean). In contrast, the short disengagement (ShortD) group included the three infants with the shortest disengagement latencies (all of them below the mean). Due to the correlations between disengagement latencies and FDs, the infants with long (or short) disengagement latencies also showed long (or short) FDs. All fixations from the three participants making up each group were combined to make a single fixation duration distribution for each viewing condition (Naturalistic and Semi-naturalistic). CRISP was then used to model these distributions.

The data from the LongD group was comprised of 486 fixations ($N_{naturalistic}$ = 274; $N_{semi\text{-}naturalistic}$ = 212). The mean fixation duration was 819.5 ms ($M_{naturalistic}$ = 778 ms, $SD_{naturalistic}$ = .152; $M_{semi\text{-}naturalistic}$ = 862 ms, $SD_{semi\text{-}naturalistic}$ = .105) and the mean disengagement latency 502 ms (SD = .137). The ShortD group included 1094 fixations ($N_{naturalistic}$ = 614; $N_{semi\text{-}naturalistic}$ = 480). The mean fixation duration was 570 ms ($M_{naturalistic}$ = 505 ms, $SD_{naturalistic}$ = .108; $M_{semi\text{-}naturalistic}$ = 635 ms, $SD_{semi\text{-}naturalistic}$ = .113) and the mean disengagement latency 87 ms (SD = .029). An

independent samples t-test showed significant differences in disengagement latencies when comparing LongD and ShortD groups ($t(4) = 5.139$, $p = .007$).

3.3.2 Modelling results

The adjustments to the model and the procedure for model fitting were the same as in simulation study 1. Figure 7 displays the empirical and simulated fixation duration distributions for the LongD (left panels) and ShortD (right panels) groups and the two viewing conditions (top: naturalistic video, bottom: semi-naturalistic videos). The reduced number of fixations for the LongD group led to rather noisy empirical fixation duration distributions (red lines). Yet it is evident from the data that, for the LongD group, the modal portion of the distribution was shifted toward longer FDs, and the tail of the distribution was somewhat increased, with a larger proportion of the data lying in the tail; this was more pronounced for the naturalistic videos than the semi-naturalistic ones. Overall, the computational model simulations were able to capture the group differences in fixation duration distributions for the LongD and ShortD groups reasonably well. Not surprisingly, the CRISP model performed worse on the more irregular LongD data as compared to the ShortD data, as is evident from the larger deviation measures (see "Error" in Table 2).

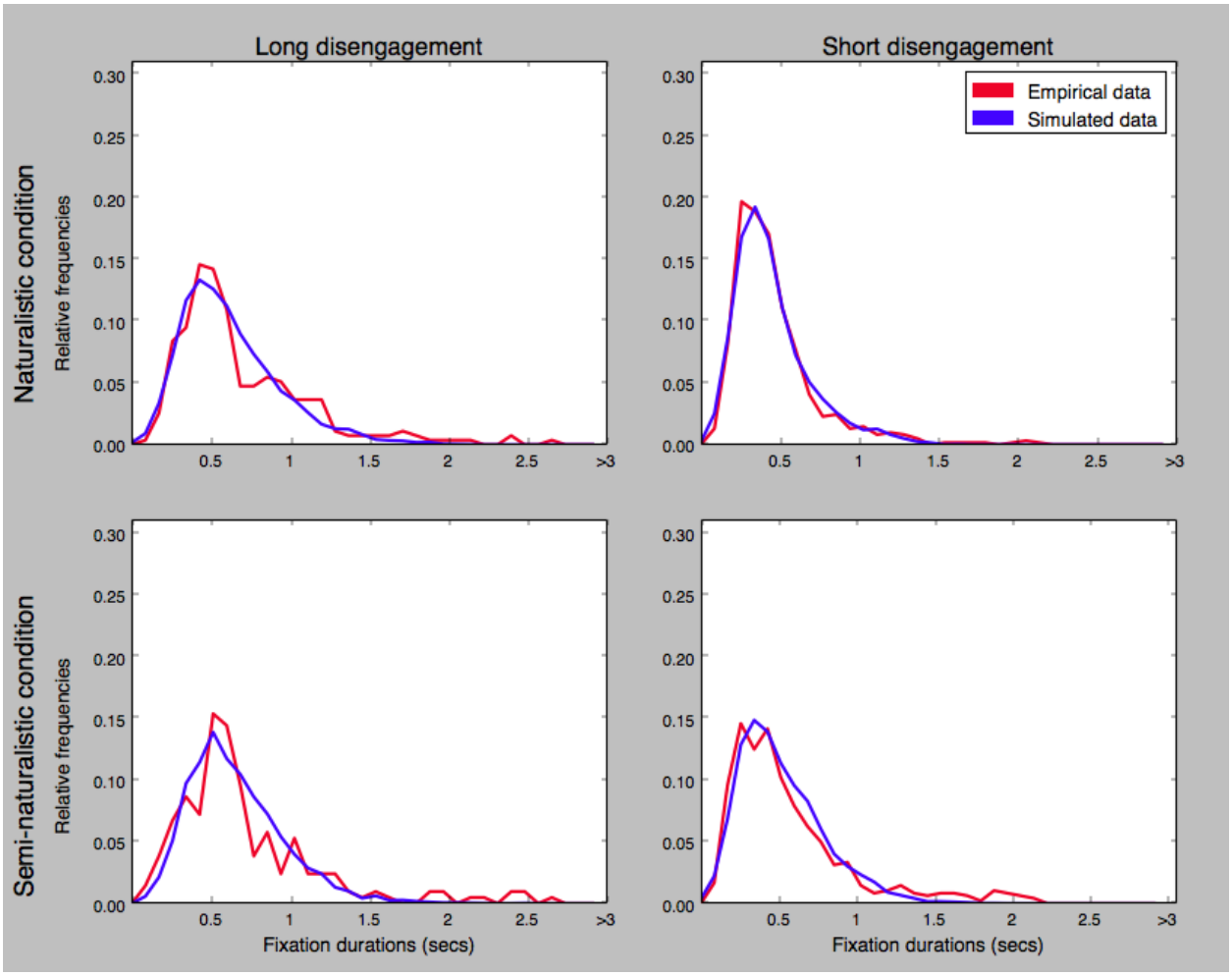

Figure 7. Simulation study 2: Simulated versus empirical fixation duration distributions in 6-month old infants. Data from long lookers (left) and short lookers (right) during viewing of naturalistic (top) and semi-naturalistic (bottom) videos.

Reducing the number of participants by grouping them according to their disengagement abilities revealed some limitations of the CRISP model in simulating infants' fixations. Whilst the peak and the tail of the distributions were efficiently captured by CRISP, there were a number of long fixations (> 2 s), particularly evident in the LongD group, that was not captured by the model. These long fixations have traditionally been associated with disengagement difficulties (see Introduction). The fact that most of these very long fixations were found in the LongD group supports this hypothesis.

Analysis of model parameters (see Table 2) offers the intriguing possibility of exploring at what level the differences in FDs between the two groups primarily originate – at the level of saccade timing, saccade programming, or both? As outlined before, we consider saccade timing parameters as an indicator for the speed of visual-cognitive processing, whereas saccade programming parameters serve as an index for the development of the oculomotor system.

The model simulations suggested a higher mean value for the random-walk timing signal for the LongD group compared to the ShortD group regardless of the viewing condition. This suggests that infants with LongD are slower at processing the visual information required for triggering a saccade program, contributing to the greater prevalence of long fixations in this group.

Notably, there are also differences in saccade programming parameters between the two groups (Table 2). In the CRISP model, at the transition from a labile to a non-labile stage of saccade programming a "point of no return" is passed. Interestingly, for the ShortD group the labile stage covers 88.95% of the total saccade programming time, while for the LongD group it covers 79.52% only. In relative terms at least, this leaves more time for the modification or cancellation of saccade programs for infants in the ShortD group, which may be indicative of a more developed oculomotor system. Furthermore, the model simulations suggested that while the non-labile stage program is already as short as for adults in the ShortD group (40 ms), it is still prolonged for the LongD group (103 ms). The neural structures involved in saccade programming (e.g., the frontal eye fields) are thought to be in place at 6 months. The present simulations are in agreement with the view that—at least in some infants—these structures may still not be developed enough to perform adult-like online control of eye movements (Csibra et al., 1998). Furthermore, in the model simulations the labile stages of saccade programming are

relatively long for both groups, which may suggest that even the ShortD group is going through a "calibration phase" that will lead to more efficient adult-like eye-movements later on. The differences between the two groups are striking, providing evidence for large variability in saccadic control across infants from the same age group.

[INSERT TABLE 2]

3.4 Qualifying remarks

A major goal of the present work was to test hypotheses related to multi-stage programming of saccades in infants. On the empirical side, we exposed 6-month-old infants to double-step trials but did not succeed in collecting reliable data. In this respect, the model simulations were used to investigate an aspect of eye-movement control that was not be tested empirically. Based on empirical findings obtained with the anti-saccade task (Johnson, 1995), we reasoned that two-stage saccade programming may be present by 4 months and then investigated more specific hypotheses with the CRISP model. Here, we take a step back and explore the more basic question whether the infants' empirical data are indeed better described by two-stage saccade programming compared with an alternative one-stage process. An affirmative answer can be derived from the model architecture along with existing simulations and features of the empirical data. One-stage programming implies that existing (labile) saccade programs cannot be cancelled. In CRISP, a saccade cancellation occurs when the timer attempts to initiate a labile saccade program while a labile program is already active. In this case, the old program is cancelled and replaced by the new program. Previous simulations using the CRISP model have revealed the role that saccade cancellation has in generating long fixation durations (Nuthmann et al., 2010). Fixation durations with no cancelations of the labile stage were shortest and had the smallest variance (Figure 4b). Average fixation duration and variance were much increased when

there was one cancellation, and even more so when there were two successive cancellations. Thus, simulations in which the cancellation mechanism is turned off (or in which one-stage programming is implemented) will generally underestimate long fixation durations (see also Trukenbrod & Engbert, 2014). Given that long tailed distributions are a prominent feature of the infant data, simulations with two-stage saccade programming are bound to fare better than without. Of course, our simulations do not "prove" that infants can cancel saccades; all we can say is that the empirical data are well fit by a model in which saccade programming progresses from a labile to a non-labile stage.

Relatedly, as explained above there are multiple routes to a longer mean fixation duration in CRISP. In the model simulations, certain model parameters were allowed to vary across age groups and/or viewing conditions, based on theoretical and empirical considerations. We then interpreted the constellation of best-fitting parameters. A complementary strategy is to run counterfactual model analyses to isolate the influence that each model component has on the mean and distribution of fixation durations (Walshe, 2015). Doing this exhaustively is neither feasible nor necessarily informative in the present context. However, in supplementary analyses (not presented here) we used this approach to follow up on the results obtained in Simulation Study 2. To determine the individual contributions of saccade programming and saccade timing model parameters in generating the fixation duration distributions from LongD and ShortD groups during free viewing of naturalistic videos, we ran simulations in which we estimated the saccade programming parameters (mean durations of labile and non-labile stages) while keeping the saccade timing parameters fixed, and vice versa. In brief, the results confirmed that, for both ShortD and LongD groups, a particular combination of saccade-programming and saccade

timing parameters was needed to achieve a good fit. Holding either set of parameters fixed did not result in an adequate fit.

4. General discussion

Though FDs and scene perception in infancy are still largely unexplored, a few studies have highlighted how FDs can be influenced by perceptual and cognitive activity (Bronson, 1990, 1994; Harris et al., 1988) and by the developmental state of the oculomotor system ("sticky fixation"; e.g., Johnson et al., 1991). It is still unclear, however, what the precise mechanisms underlying saccadic control in infancy are and the extent to which large variations in FD distributions found during the first months of life are affected by these factors. Investigating these issues using more traditional psychophysical experiments can be challenging, if not unworkable, due to the practical and current technological limitations that testing infants entails. On the other hand, computational modeling allows us to describe, predict and explain data that is itself unobservable (Lewandowsky & Farrell, 2011) and complements more traditional methods by investigating the mechanisms that are not directly accessible through experimentation (Braitenberg, 1984; Schlesinger & McMurray, 2012).

The present work aimed at advancing our understanding of the mechanisms underlying eye-movement control in infancy. To this end, we reported an empirical study and two simulation studies using the CRISP model of FDs in (adult) scene viewing (Nuthmann et al., 2010). The goal of a first baseline simulation was to test the CRISP model's generalizability to fixation-duration data from 6-month-old infants. In a second simulation study, we investigated differences in fixation durations in infants with long and short disengagement latencies.

For the empirical study we recorded eye-movements from groups of 6-month-old infants and adults while watching customized naturalistic and semi-naturalistic videos. Participants also

performed a gap-overlap task (e.g., Hood & Atkinson, 1993) to measure their disengagement abilities. This design allowed us to investigate the effect of viewing condition on fixation duration distributions and to examine the relationship between disengagement abilities and fixation durations. The results showed that mean fixation durations were significantly longer for 6-month-olds than for adults, and longer for semi-naturalistic compared with naturalistic videos, with no interaction between age group and viewing condition. In addition, for infants we found high correlations between disengagement latencies and FDs in both viewing conditions. In infants, saccade execution is greatly affected by the development of the frontal eye fields and other neural structures. Thus, infants showing more difficulties disengaging are thought to have a less developed oculomotor control system, producing longer mean FDs. However, this developmental argument cannot explain the correlation that we found, though reduced, in adults when they viewed naturalistic videos. Taken together, the results suggest that disengagement is not only a consequence of an underdeveloped oculomotor system, but that it also affects saccade execution as a function of the characteristics of the visual stimuli being processed. Consequently, disengagement could be used to quantify the stimulus dependency of FDs. These results fit with the findings from a study by Kikuchi and colleagues (2011) who found larger disengagement latencies and saccade-related event-related potentials (ERPs) when children disengaged from faces, suggesting that the encoding and processing of the foveated stimulus plays a role in the ability to shift the gaze from a central target.

The results from simulation study 1 showed that the CRISP model was able to capture the trends observed in empirical fixation-duration data from 6 month-old infants. A key assumption of the model is that saccades are programmed in two discrete stages—an initial labile stage that is subject to cancelation if another saccade program is subsequently initiated, followed by a non-

labile stage that cannot be cancelled. This distinction is based on results from the double-step paradigm (Becker & Jürgens, 1979; Walshe & Nuthmann, 2015). To gather sufficient and interpretable double-step data from 6-month-olds proved to be a very challenging task in the present study. Here, we reported computer simulations which showed that infants' fixation duration distributions were well captured by a model implementing two-stage saccade programming. From these results, one may tentatively conclude that two-saccade programming is active at this age. A comparison of the parameters estimated for the different age groups suggests that both saccade timing and saccade programming (labile and non-labile stages) are considerably slowed down at 6 months compared to adults. This may suggest that at 6 months, fixation durations are still affected by the development of the oculomotor system (indicated by saccade programming parameters) but also by the efficiency of visual and cognitive processing (indicated by saccade timing parameters).

Simulation study 2 aimed to explore a developmentally important research question: how do long and short lookers differ in terms of the maturity of their oculomotor and visual/cognitive processing? This study acts as a first demonstration of how CRISP could be used to investigate questions about the development of active vision that are not feasible given the limitations of infant behavioral methodologies. The differences in FDs between infants with large (LongD group) and short (ShortD group) disengagement latencies were modeled using CRISP. Based on previous research on the inhibition of saccades in infants (e.g., Johnson, 1995), we reasoned that infants with larger disengagement difficulties may have longer saccade latencies, with both the labile and non-labile phases being prolonged. The model simulations revealed differences between the ShortD and LongD groups in both saccade timing as well as saccade programming parameters. According to the model architecture, this means that FDs at 6 months are affected

both by the speed of visual-cognitive processing as well as developmental aspects of the oculomotor system, with the latter being particularly prominent in infants with disengagement difficulties. These results are in line with findings from Domsch and colleges (2010), reporting that looking behavior within a habituation paradigm was affected by both information processing and disengagement.

We acknowledge that there is a risk of over-interpreting the results from simulation studies like the ones presented here. Ideally, CRISP should be used to generate hypotheses that are then empirically validated, rather than the simulation be considered the end goal. We now comment on the interpretation of saccade timing parameters in the model (i.e., the mean and variance of the timing signal, which describe the random-walk transition rate N/ t_{sac}). In some of our considerations, we have discussed them as an indicator for the difficulty or speed of visual-cognitive processing. In the present context, this is little more than a "black box" as our present work was concerned with differences in overall fixation duration distributions between infants and adults, and with the global effects of the viewing condition, not the local effects of the fixated information. Empirical and computational research on fixation durations in adults has begun to elucidate the CRISP model's key assumption that moment-to-moment difficulties in visual and cognitive processing can immediately inhibit (i.e., delay) saccade initiation, leading to longer fixation durations. One question concerns the degree to which fixation durations are under the direct moment-to-moment control of the current visual stimulus. Previous simulations with the CRISP model have investigated this issue by applying the scene onset delay paradigm, which selectively manipulates global scene processing difficulty (Henderson & Pierce, 2008; Henderson & Smith, 2009). The empirical data and model simulations showed that fixation durations are, at least partly, influenced by the current visual stimulus (Nuthmann et al., 2010).

Going beyond global effects, Nuthmann (2016) investigated properties of local control of fixation durations—that is, whether fixation duration varies as a function of the processing difficulty of the currently foveated scene content. Linear mixed models were used to assess the independent effects of five low-level, mid-level, and higher-level visual image features, yielding local image-based indexes of processing difficulty. For example, as the number of edges in foveal vision increased, fixation duration increased as well. Interestingly, modelling such data with CRISP (or a related model) would more fully elucidate the control mechanisms related to the random-walk saccade timer: The transition rate of the random-walk can locally adjust to the visual-cognitive complexity of the currently fixated scene region. Of course, such an endeavor would require us to add a "where" module to the model, representing the spatial decision about where to fixate next. Clearly, more empirical and computational research is needed to unravel visual-cognitive processing during scene viewing, by isolating the precise contributions of low-level bottom-up and high-level top-down factors to fixation durations in infants, and to investigate the general principles of the interaction between visual-cognitive processing and saccade timing in both infants and adults.

Simulation study 2 revealed some limitations of the CRISP model when simulating FDs from infants. In adults, very long FDs (> 2 s) are fairly uncommon and tend to be discarded from the analysis, the rationale being that such long fixations may not be indicative of any perceptual activity or information processing (Inhoff & Radach, 1998). Young infants, however, can present a number of very long fixations that are mainly assumed to be a consequence of disengagement difficulties. For this reason we decided not to discard long fixations from data analysis. The model simulations were able to capture the trends in fixation duration distributions as expressed in the mode and the slope of the distribution. Perhaps not suprisingly, they did not capture the

very long fixations present in the tail of the distribution. The fact that this problem was more evident in the LongD group may suggest that these fixations were a consequence of poor disengagement abilities. To accommodate this, in future work the model may be extended by including a "disengagement component" related to saccade programming. The disengagement component will stocastically prolong certain fixations, with the result that the simulations may recover the very long fixations found in infants with disengagement difficulties. In this context, future research may explore whether disengagement selectively prolongs the non-labile stage of saccade programming only, the labile stage as well, or whether disengagement is actually related to the transition between the two stages or from non-labile to saccade execution. Further empirical work is required.

Whilst the present studies helped to shed light on the mechanisms and processes underlying eye-movements in 6-month-old infants, the causes that explain the large individual differences between short and long lookers are still unknown. For instance, from a skill learning perspective, it could be argued that infants with shorter FDs will generate a larger number of saccades per day, as a consequence of which they will become more skilled in this aspect of motor control compared with other infants. In order to test this hypothesis, future simulations may consider factoring a "learning component" into the model.

Finally, in the present article we investigated the stability in FDs within a single testing session at only one infant age. Existing evidence suggests that newborn and very young infant saccade programming may be very primitive and involve relatively fixed amplitude saccades with limited scope for on-line modification (Aslin & Salapatek, 1975; Johnson, 1990), which is suggestive of one-stage saccade programming. Such evidence must be supported by future precise behavioral studies analyzing the timing of infant saccades and further computational

modeling of fixation durations for infants younger than 6 months. Similarly, future longitudinal studies should explore the trajectories of the same infants at different points in time in order to investigate typical developmental of oculomotor control and provide a reference for atypical development such as in the case of Autism Spectrum Disorder or Attention Deficit and Hyperactivity Disorder.

Acknowledgments

The authors wish to thank Mats B. Küssner for his comments on the first draft of this manuscript and R. Calen Walshe for his assistance during this project. Thanks also to Casey Thornton for performing the hand-coding reliability analysis. This research was supported by the EC Marie Curie Initial Training Networks FP7-PEOPLE-2010-ITN PART B, and the UK Medical Research Council.

References

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: a reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 997–1010. http://doi.org/10.1037/0278-7393.25.4.997

Aslin, R. N., & Salapatek, P. (1975). Saccadic localization of visual targets by the very young human infant. *Perception & Psychophysics*, *17*(3), 293–302. http://doi.org/10.3758/BF03203214

Atkinson, J. (2000). *The Developing Visual Brain*. Oxford: Oxford University Press.

Atkinson, J., Hood, B., Wattam-Bell, J., & Braddick, O. (1992). Changes in infants' ability to switch visual attention in the first three months of life. *Perception*, *21*, 643–653. http://doi.org/10.1068/p210643

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, *19*(9), 967–983. http://doi.org/10.1016/0042-6989(79)90222-0

Blignaut, P., & Wium, D. (2014). Eye-tracking data quality as affected by ethnicity and experimental design. *Behavior Research Methods*, *46*(1), 67–80. http://doi.org/10.3758/s13428-013-0343-0

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 185–207. http://doi.org/10.1109/tpami.2012.89

Braddick, O., Atkinson, J., Hood, B., Harkness, W., Jackson, G., & Vargha-Khademt, F. (1992). Possible blindsight in infants lacking one cerebral hemisphere. *Nature*, *360*, 461–463. http://doi.org/10.1038/360461a0

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. http://doi.org/10.1163/156856897X00357

Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology*. Cambridge: MIT press.

Bronson, G. W. (1974). The postnatal growth of visual capacity. *Child Development*, *45*(4), 873–890. http://doi.org/10.2307/1128073

Bronson, G. W. (1990). Changes in infants' visual scanning across the 2-to 14-week age period. *Journal of Experimental Child Psychology*, *49*(1), 101–125. http://doi.org/10.1016/0022-0965(90)90051-9

Bronson, G. W. (1994). Infants' transitions toward adult-like scanning. *Child Development*, *65*(5), 1243–1261. http://doi.org/10.1111/j.1467-8624.1994.tb00815.x

Brown, J. H., Johnson, M. H., Paterson, S. J., Gilmore, R., Longhi, E., & Karmiloff-Smith, A. (2003). Spatial representation and attention in toddlers with Williams syndrome and Down syndrome. *Neuropsychologia*, *41*(8), 1037–1046. http://doi.org/10.1016/S0028-3932(02)00299-3

Butcher, P. R., Kalverboer, A. F., & Geuze, R. H. (2000). Infants' shifts of gaze from a central to a peripheral stimulus: A longitudinal study of development between 6 and 26 weeks. *Infant Behavior and Development*, *23*(1), 3–21. http://doi.org/10.1016/s0163-6383(00)00031-x

Camalier, C. R., Gotler, A., Murthy, A., Thompson, K. G., Logan, G. D., Palmeri, T. J., & Schall, J. D. (2007). Dynamics of saccade target selection: Race model analysis of double step and search step saccade production in human and macaque. *Vision Research*, *47*(16), 2187–2211. http://doi.org/10.1016/j.visres.2007.04.021

Canfield, R. L., Wilken, J., Schmerl, L., & Smith, E. G. (1995). Age-related change and stability of individual differences in infant saccade reaction time. *Infant Behavior and Development*, *18*(3), 351–358. http://doi.org/10.1016/0163-6383(95)90023-3

Carlson, E., & Triesch, J. (2004). A computational model of the emergence of gaze following. *Progress in Neural Processing*, *15*, 105–114. http://doi.org/10.1142/9789812702784_0010

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9(3):6*, 1–15. http://doi.org/10.1167/9.3.6

Csibra, G., Tucker, L. A., & Johnson, M. H. (1998). Neural correlates of saccade planning in infants: A high-density ERP study. *International Journal of Psychophysiology*, *29*(2), 201–215. http://doi.org/10.1016/s0167-8760(98)00016-6

Csibra, G., Tucker, L. A., & Johnson, M. H. (2001). Differential frontal cortex activation before anticipatory and reactive saccades in infants. *Infancy*, *2*(2), 159–174. http://doi.org/10.1207/s15327078in0202_3

Domsch, H., Lohaus, A., & Thomas, H. (2010). Influences of information processing and disengagement in infants' looking behaviour. *Infant and Child Development*, *19*(2), 161–174. http://doi.org/10.1002/icd.647

Elsabbagh, M., Volein, A., Holmboe, K., Tucker, L., Csibra, G., Baron-Cohen, S., … Johnson, M. H. (2009). Visual orienting in the early broader autism phenotype: disengagement and facilitation. *Journal of Child Psychology and Psychiatry*, *50*(5), 637–642. http://doi.org/10.1111/j.1469-7610.2008.02051.x

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*(5), 621–636. http://doi.org/10.1016/S0042-6989(01)00301-7

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813. http://doi.org/10.1037/0033-295X.112.4.777

Farroni, T., Simion, F., Umiltà, C., & Barba, B. D. (1999). The gap effect in newborns. *Developmental Science*, *2*(2), 174–186. http://doi.org/10.1111/1467-7687.00066

Findlay, J. M., & Harris, L. R. (1984). Small saccades to double-stepped targets moving in two dimensions. In A. G. Gale & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research* (pp. 71–78). Amsterdam: Elsevier. http://doi.org/10.1016/S0166-4115(08)61820-8

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences*, *16*(3), 553–567. http://doi.org/10.1017/S0140525X00031575

Frick, J. E., Colombo, J., & Saxon, T. F. (1999). Individual and developmental differences in disengagement of fixation in early infancy. *Child Development*, *70*(3), 537–48. http://doi.org/10.1111/1467-8624.00039

Gilmore, R. O., & Johnson, M. H. (1997a). Body-centered representations for visually-guided action emerge during early infancy. *Cognition*, *65*(1), B1–B9. http://doi.org/10.1016/S0010-0277(97)00038-3

Gilmore, R. O., & Johnson, M. H. (1997b). Egocentric action in early infancy: Spatial frames of reference for saccades. *Psychological Science*, *8*(3), 224–230. http://doi.org/10.1111/j.1467-9280.1997.tb00416.x

Hainline, L., Turkel, J., Abramov, I., Lemerise, E., & Harris, C. M. (1984). Characteristics of saccades in human infants. *Vision Research*, *24*(12), 1771–1780. http://doi.org/10.1016/0042-6989(84)90008-7

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34.

Harris, C. M., Hainline, L., Abramov, I., Lemerise, E., & Camenzuli, C. (1988). The distribution of fixation durations in infants and naive adults. *Vision Research*, *28*(3), 419–432. http://doi.org/10.1016/0042-6989(88)90184-8

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of flxation durations. *Psychonomic Bulletin & Review*, *15*(3), 566–573. http://doi.org/10.3758/PBR.15.3.566

Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition*, *17*(6–7), 1055–1082. http://doi.org/10.1080/13506280802685552

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228. http://doi.org/10.1037//0096-1523.25.1.210

Hessels, R. S., Andersson, R., Hooge, I. T. C., Nyström, M., & Kemner, C. (2015). Consequences of eye color, positioning, and head movement for eye-tracking data quality in infant research. *Infancy*, *20*(6), 601–633. http://doi.org/10.1111/infa.12093

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye Tracking. A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: What it is and how to measure it. In *Proceedings of the 2012 Symposium on Eye Tracking Research and Applications* (pp. 45–52). ACM. http://doi.org/10.1145/2168556.2168563

Hood, B. M., & Atkinson, J. (1990). Sensory visual loss and cognitive deficits in the selective attentional system of normal infants and neurologically impaired children. *Developmental Medicine & Child Neurology*, *32*(12), 1067–1077. http://doi.org/10.1111/j.1469-8749.1990.tb08525.x

Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Infant Behavior and Development*, *16*(4), 405–422. http://doi.org/10.1016/0163-6383(93)80001-o

Hunnius, S., & Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy*, *6*(2), 231–255. http://doi.org/10.1207/s15327078in0602_5

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–53). Oxford: Elsevier Science Ltd. http://doi.org/10.1016/B978-008043361-5/50003-1

Johnson, M. H. (1990). Cortical maturation and the development of visual attention in early infancy. *Journal of Cognitive Neuroscience*, *2*(2), 81–95. http://doi.org/10.1162/jocn.1990.2.2.81

Johnson, M. H. (1995). The inhibition of automatic saccades in early infancy. *Developmental Psychobiology*, *28*(5), 281–291. http://doi.org/10.1002/dev.420280504

Johnson, M. H. (2011). *Developmental cognitive Neuroscience* (Third edit). Oxford: Wiley-Blackwell.

Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1991). Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience*, *3*(4), 335–344. http://doi.org/10.1162/jocn.1991.3.4.335

Kikuchi, Y., Senju, A., Akechi, H., Tojo, Y., Osanai, H., & Hasegawa, T. (2011). Atypical disengagement from faces and its modulation by the control of eye fixation in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *41*(5), 629–645. http://doi.org/10.1007/s10803-010-1082-z

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 14.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288. http://doi.org/10.1016/s0010-0277(87)80006-9

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. Thousand Oaks. CA: Sage.

Loftus, G. R. (1985). Picture perception: Effects of luminance on available information and information-extraction rate. *Journal of Experimental Psychology: General*, *114*(3), 342–356. http://doi.org/10.1037/0096-3445.114.3.342

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 565–572. http://doi.org/10.1037//0096-1523.4.4.565

Mareschal, D. (2010). Computational perspectives on cognitive development. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 696–708. http://doi.org/10.1002/wcs.67

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*(12), 899–917. http://doi.org/10.1037/h0037368

Matsuzawa, M., & Shimojo, S. (1997). Infants' fast saccades in the gap paradigm and development of visual attention. *Infant Behavior and Development*, *20*(4), 449–455. http://doi.org/10.1016/s0163-6383(97)90035-7

McAuley, J. H., Rothwell, J. C., & Marsden, C. D. (1999). Human anticipatory eye movements may reflect rhythmic central nervous activity. *Neuroscience*, *94*(2), 339–350. http://doi.org/10.1016/S0306-4522(99)00337-1

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586. http://doi.org/10.3758/BF03203972

Nuthmann, A. (2016). Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic Bulletin & Review*, 1–23. http://doi.org/10.3758/s13423-016-1124-4

Nuthmann, A., & Engbert, R. (2009). Mindless reading revisited: An analysis based on the SWIFT model of eye-movement control. *Vision Research*, *49*(3), 322–336. http://doi.org/10.1016/j.visres.2008.10.022

Nuthmann, A., & Henderson, J. M. (2012). Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism. *Visual Cognition*, *20*(4–5), 457–494. http://doi.org/10.1080/13506285.2012.670142

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382–405. http://doi.org/10.1037/a0018924

Nyström, M., Andersson, R., Holmqvist, K., & van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, *45*(1), 272–288. http://doi.org/10.3758/s13428-012-0247-4

Papageorgiou, K. A., Farroni, T., Johnson, M. H., Smith, T. J., & Ronald, A. (2015). Individual differences in newborn visual attention associate with temperament and behavioral difficulties in later childhood. *Scientific Reports*, *5*. http://doi.org/10.1038/srep11264

Papageorgiou, K. A., Smith, T. J., Wu, R., Johnson, M. H., Kirkham, N. Z., & Ronald, A. (2014). Individual differences in infant fixation duration relate to attention and behavioral control in

childhood. *Psychological Science*, *25*(7), 1371–1379. http://doi.org/10.1177/0956797614531295

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. http://doi.org/10.1037//0033-2909.124.3.372

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2013). Using EZ Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review*, *33*(2), 110–149. http://doi.org/10.1016/j.dr.2013.03.001

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125–157. http://doi.org/10.1037/0033-295X.105.1.125

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*(4), 445–476. http://doi.org/10.1017/S0140525X03000104

Robertson, S. S. (1982). Intrinsic temporal patterning in the spontaneous movement of awake neonates. *Child Development*, *53*(4), 1016–1021. http://doi.org/10.2307/1129142x

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Caro, D. M. (2002). A longitudinal study of visual expectation and reaction time in the first year of life. *Child Development*, *73*(1), 47–61. http://doi.org/10.1111/1467-8624.00391

Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, *24*(2), 113–121. http://doi.org/10.1016/S0166-2236(00)01685-4

Rothbart, M. K., Posner, M. I., & Rosicky, J. (1994). Orienting in normal and pathological development. *Development and Psychopathology*, *6*(4), 635–652. http://doi.org/10.1017/s0954579400004715

Saez de Urabain, I. R., Johnson, M. H., & Smith, T. J. (2015). GraFIX: A semiautomatic approach for parsing low- and high-quality eye-tracking data. *Behavior Research Methods*, *47*(1), 53–72. http://doi.org/10.3758/s13428-014-0456-0

Schlesinger, M., Amso, D., & Johnson, S. P. (2007). The neural basis for visual selective attention in young infants: A computational account. *Adaptive Behavior*, *15*(2), 135–148. http://doi.org/10.1177/1059712307078661

Schlesinger, M., & McMurray, B. (2012). The past, present, and future of computational models of cognitive development. *Cognitive Development*, *27*(4), 326–348. http://doi.org/10.1016/j.cogdev.2012.07.002

Sirois, S., & Mareschal, D. (2002). Models of habituation in infancy. *Trends in Cognitive Sciences*, *6*(7), 293–298. http://doi.org/10.1016/s1364-6613(02)01926-5

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*(1), 29–56. http://doi.org/10.1207/s15516709cog1401_3

Trukenbrod, H. A., & Engbert, R. (2014). ICAT: a computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review*, *21*(4), 907–934. http://doi.org/10.3758/s13423-013-0575-0

Walshe, R. C. (2015). *The operation of eye-movement control mechanisms during the perception of naturalistic scenes (Unpublished doctoral thesis)*. University of Edinburgh, Edinburgh, UK.

Walshe, R. C., & Nuthmann, A. (2014). Asymmetrical control of fixation durations in scene viewing. *Vision Research*, *100*, 38–46. http://doi.org/10.1016/j.visres.2014.03.012

Walshe, R. C., & Nuthmann, A. (2015). Mechanisms of saccadic decision making while encoding naturalistic scenes. *Journal of Vision*, *15(5):21*, 1–19. http://doi.org/10.1167/15.5.21

Wass, S. V, Porayska-Pomsta, K., & Johnson, M. H. (2011). Training attentional control in infancy. *Current Biology*, *21*(18), 1543–1547. http://doi.org/10.1016/j.cub.2011.08.004

Wass, S. V, Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, *45*(1), 229–250. http://doi.org/10.3758/s13428-012-0245-6

Westheimer, G. (1954). Eye movement responses to a horizontally moving visual stimulus. *AMA Archives of Ophthalmology*, *52*(6), 932–941.

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R., & Vogeley, K. (2010). It's in your eyes--using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience*, *5*(1), 98–107. http://doi.org/10.1093/scan/nsq024

Wolff, P. H. (1991). Endogenous motor rhythms in young infants. In J. Fagard & P. H. Wolff (Eds.), *The development of timing control and temporal organization in coordinated action* (pp. 119–133). Amsterdam: North Holland. http://doi.org/10.1016/S0166-4115(08)60762-1

Wu, C.-C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, *5*, 54. http://doi.org/10.3389/fpsyg.2014.00054

Appendix A. Fixation detection and coding

The quality of the raw data generated by the eye-tracker may vary depending on many different factors such as the eye-tracker model and manufacturer, the eye physiology, the calibration procedure, the position of the participant relative to the eye-tracker, the degree of head, or even ethnicity (Blignaut & Wium, 2014; Holmqvist et al., 2011; Holmqvist, Nyström, & Mulvey, 2012; Nyström, Andersson, Holmqvist, & van de Weijer, 2013; Saez de Urabain et al., 2015). Eye-tracking data coming from populations such as infants may contain considerably higher levels of noise than data from more compliant participants. This is not only due to their high degree of movement, but also a consequence of poor calibration procedures and issues related to corneal reflection techniques that affect the youngest ages (Saez de Urabain et al., 2015, for review).

Fixations were detected using GraFIX, a semi-automatic approach for parsing low- and high-quality eye-tracking data (Saez de Urabain et al., 2015). The input is initially parsed by using velocity-based algorithms whose input parameters are adapted by the user, and then manipulated using the graphical interface, allowing accurate and rapid adjustments of the algorithms' outcome. Saez de Urabain and colleagues (2015) validated this technique and demonstrated its suitability for infant data.

The input parameters for the automatic detection algorithms were adapted according to the data quality for each participant and visit, having two different sets of parameters for participants featuring lower (>0.25° root mean square of inter-sample distances, RMS) or higher (<0.25° RMS) data quality (see Table 3). These parameters, chosen after executing the algorithms with a wide range of values and evaluating the outcomes, may not necessarily be optimal in data sets with other characteristics. Having only two sets of parameters rather than multiple options can

facilitate the coding process, especially for novice users (Saez de Urabain et al., 2015). In order to keep the same standards across participants and coders strict coding guidelines were defined. A fixation was coded when both the x and y coordinates were stable at one point, i.e. when the plots of both x and y coordinates against time were displaying horizontal lines. If the detection of one eye was imprecise the data from the other eye was used. If the coder was not entirely sure about coding a particular fixation they were advised to leave it out. Saccades that were too short to be detected by the algorithms were also coded. Fixations that were cut by blinks and smooth pursuit eye-movements (diagonal movement of the X/Y trace) were deleted.

Appendix B. Details on fitting model parameters

The best set of parameters from the genetic algorithm was selected as follows. First, the deviation between the two distributions was measured by calculating the mean-root-square error. For each viewing condition, fixation duration distributions were calculated from 46 bins with bin centers ranging from 80 to 6000 ms, in steps of 130 ms. Secondly, the difference between the means was evaluated by the mean-squared normalized error for average FDs (Reichle et al., 1998). As a result of the long tail and the high skewedness of the fixation duration distributions presented in this paper, the simulation with the lower E did not necessarily indicate the best fit. To be a good fit, a simulation needs to capture the peak and the tail of the distribution. The fixation duration distributions reported in this paper include particularly long fixations (> 2 s) , which are typically excluded from the analysis when studying adults (e.g., Nuthmann et al., 2010). These long fixations, however, are not uncommon in young infants and can in fact be revealing when studying the development of oculomotor control. We decided to include these fixations for both age groups in order to maintain consistency in the analysis, despite the additional difficulties this entails when fitting the data with the genetic algorithm. For instance, when the empirical distribution has a long tail, the simulation that reproduces the peak and the tail of the distribution best is not necessarily the one that shows the lowest E: mean FDs for the empirical and simulated data can still be far apart, which increases the E value. Occasionally, this will affect the parameter estimation from the genetic algorithm. To overcome this issue, we selected the set of parameters that (1) fitted the peak and the tail of the empirical distribution, and (2) presented the best goodness-of-fit measure.

Tables

Table 1 Best-fitting parameters for simulation study 1

	Parameter	Symbol	Range	Naturalistic	Semi-naturalistic
Infants	Saccade programming				
	Labile stage (ms)	τ_{lab}	200-400		385
	Non-labile stage (ms)	τ_{nlab}	30-120		62
	Saccade timing				
	Mean (ms)	t_{sac}	200-650	458	496
	Variance	N	5-20	20	5
	Saccade execution (ms)	τ_{ex}	---		50
	Standard deviation of gamma distributions	σ_γ	---		0.33
	Error	E	---	4.275	3.718
	Error cumulative distribution	E_c	---	4.093	9.821
Adults	Saccade programming				
	Labile stage (ms)	τ_{lab}	150-350		224
	Non-labile stage (ms)	τ_{nlab}	30-50		37
	Saccade timing				
	Mean (ms)	t_{sac}	150-650	253	271
	Variance	N	5-20	20	10
	Saccade execution (ms)	τ_{ex}	---		37
	Standard deviation of gamma distributions	σ_γ	---		0.25
	Error	E	---	4.513	6.042
	Error cumulative distribution	E_c	---	6.758	4.248

Table 2 Best-fitting parameters for simulation study 2

Parameter	Symbol	Range	LongD		ShortD	
			Nat	Non-nat	Nat	Non-nat
Saccade Programming						
Labile stage (ms)	τ_{lab}	200-400	400		322	
Non-labile stage (ms)	τ_{nlab}	30-120	103		40	
Saccade timing						
Mean (ms)	t_{sac}	200-650	560	648	437	466
Variance	N	5-20	10	10	20	6
Saccade execution (ms)	τ_{ex}	---	50			
Standard deviation of gamma distributions	σ_γ	---	0.33			
Error	E	---	14.641	23.684	11.643	12.747
Error cumulative distribution	E_c	---	4.244	8.680	4.950	3.715

Table 3 Input Values for the automatic detection algorithms in low and high quality data sets.

	High spatial precision	Low spatial precision
Interpolation latency (ms)	60	60
Velocity threshold (°/sec)	9	20
Maximum interpolation displacement (°)	0.25	0.25
Degree per pixel (°/pix)	0.0177	0.0177
Maximum distance for merging adjacent fixations (°)	0.24	0.35
Maximum time for merging adjacent fixations (ms)	50	50
Maximum RMS per fixation (°)	0.24	0.21
Minimum fixation duration (ms)	99	120